



Fernando, WAC., Canagarajah, CN., & Bull, DR. (1999). Sudden scene change detection in MPEG-2 video sequences. In *Proc. IEEE International Workshop on Multimedia Signal Processings* (pp. 259 - 264). Institute of Electrical and Electronics Engineers (IEEE).  
<https://doi.org/10.1109/MMSP.1999.793842>

Peer reviewed version

Link to published version (if available):  
[10.1109/MMSP.1999.793842](https://doi.org/10.1109/MMSP.1999.793842)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

## SUDDEN SCENE CHANGE DETECTION IN MPEG-2 VIDEO SEQUENCES

W.A.C. Fernando, C.N. Canagarajah, D. R. Bull

Image Communications Group, Centre for Communications Research, University of Bristol,  
Merchant Ventures Building, Woodland Road, Bristol BS8 1UB, United Kingdom.

Voice - +44-117-954-5198, Fax - +44-117-954-5206

Email - W.A.C.Fernando@bristol.ac.uk

**Abstract** - Content based indexing and retrieval of video is becoming increasingly important in many applications. Identifying scene changes and special effects in a video scene is an essential pre-requisite for automatic indexing. This paper presents a real time algorithm, which can detect abrupt scene changes in compressed domain. It is based on the number of interpolated macroblocks (MBs) for a given B-frame as the main feature since it expresses a measure of how strong the previous and future I or P (I/P) frames are correlated. Experimental results show that this algorithm can detect most abrupt scene changes in MPEG-2 compressed video.

### I. INTRODUCTION

Advances in communication technology have led to proliferation of multimedia information systems that allow users to store, retrieve and manipulate different types of information. There are several applications including distance learning, digital libraries, telemedicine, interactive television and multimedia news, which are expected to involve visual information systems. MPEG-2 video compression [1] is used in many current and emerging products. It is at the heart of digital television set-top boxes, HDTV decoders, video conferencing, internet video and other applications. These applications benefit from video compression in the fact that they may require less storage space for archived video information, less bandwidth for the transmission of the video information from one point to the other, or a combination of both. The main feature required in a visual information system is an efficient video indexing technique to enable fast access of the stored data. A video index serves as a descriptor of the video, thus enabling fast access to the video clips stored in a multimedia database. It is essential to segment the video into elemental scenes to aid automatic indexing.

Video is a structured medium and thus can be decomposed into scenes, shots and frames. A frame is a basic unit of the video and corresponds to a signal image. Shots are sequence of frames recorded successively by one camera. Scenes are usually composed of a small number of interrelated shots that are unified by location. Figure 1 shows this hierarchy. Since frames belonging to two successive shots are likely to be visually dissimilar, techniques for video parsing rely on detecting points in the sequence at which a quantitative measure expressing the difference between successive frames exceeds a certain threshold. These points are known as shot transitions. Transitions can be divided into two categories: sudden (abrupt) transitions and gradual transitions. Several algorithms are proposed for sudden and gradual scene change detection [2-7].

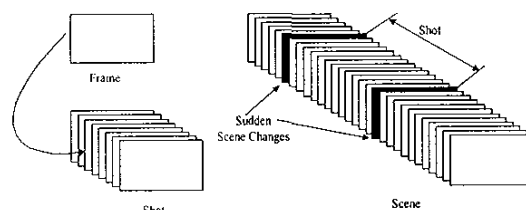


Figure 1: A Hierarchical description of video sequence

Yeo proposed DC sequences for sudden change detection [2] in MPEG-2 compressed video. This method leads to partial decompression of video and hence it is costly. Zhang [3] and Shin [4] used forward and backward predicted information to detect abrupt scene changes. This algorithm is very sensitive to the thresholds and the type of sequences. It detected many false scene changes when the scene contains special effects. Furthermore, the thresholds are heavily dependent on the nature of the sequences. This paper presents a simple algorithm based on the number of interpolated MBs in B-frames of MPEG-2 compressed video.

Rest of the paper is organised as follows. Section II presents an overview of MPEG-2. Complete algorithm is given in section III and section IV includes some simulation results. Finally, section V presents the conclusions.

## II. MPEG-2 OVERVIEW

MPEG-2 video compression is used in many current and emerging products. MPEG-2 video is broken up into a hierarchy of layers to help with error handling, random search editing, and synchronisation. From the top level, the first layer is known as the video sequence layer. The second layer down is the group of pictures (GOP), which is composed of one or more groups of intra (I) frames and/or non-intra (P and/or B) frames. The third layer down is the picture layer itself, and the next layer beneath it is called the slice layer. Each slice consists of MBs, which are 16x16 arrays of luminance pixels, or picture data elements, with 8x8 arrays of associated chrominance pixels. The MBs can further be divided into 8x8 blocks, for further processing.

### MPEG-2 Picture Types

**Intra Pictures (I-Pictures)** These pictures are encoded only with respect to themselves. Here each picture is decomposed into blocks of 8x8 pixels each and encoded directly using DCT transformation process.

**Predictive Pictures (P-Pictures)** These are pictures encoded using motion compensated prediction from a past I/P picture. A prediction error is calculated between a 16x16 pixels region (MB) in the current picture and the past reference I/P picture. A motion vector is also calculated to determine the value and direction of the prediction. The prediction error is transmitted after DCT coding.

**Bi-directional Pictures (B-Pictures)** These are pictures encoded using motion compensation prediction from a past and/or next I/P picture. A prediction error is calculated between a 16x16 pixels region in the current picture and the past as well as next reference I/P picture. Two motion vectors are also calculated to determine the value and direction of the prediction from both directions.

Figure 2 shows a typical MPEG-2 video sequence with GOP of 12 and sub-GOP size of 3.

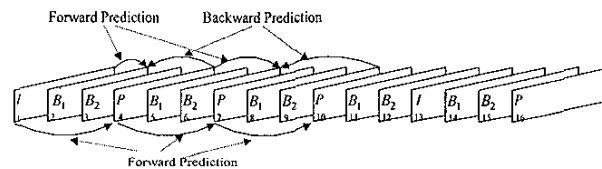


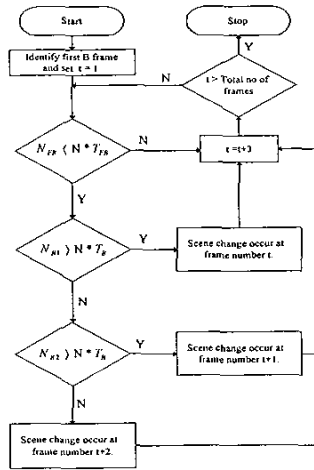
Figure 2 : A Typical MPEG-2 Compressed Video Sequence

### III. PROPOSED ALGORITHM

B-frames are encoded using motion compensated predictions from a past and/or next I/P frame. Therefore, B-frames should have three types of MBs (rarely, a few MBs are intra-coded) i.e.

1. Forward predicted MBs (predicted from previous I/P frame)
2. Backward predicted MBs (predicted from next I/P frame)
3. Interpolated MBs (predicted from both previous and next I/P frame)

Thus, the number of interpolated MBs ( $N_{FB}$ ) for a given B-frame expresses a measure of how strong the correlation between the previous and next I/P frames. If the number of interpolated MBs for a given first B-frame is high (above  $N * T_{FB}$ , where,  $T_{FB}$  - pre-determined percentage threshold and  $N$  - total number of MBs in a frame), and then there is a strong correlation between previous and next I/P frame with respect to the current B-frame. Therefore, it is not possible to have an abrupt scene change between these two I/P frames in the vicinity of the current B-frame. If it is below the threshold, that indicates that the previous and next I/P frames are not correlated. This situation leads to an abrupt scene change either at one of the B-frames or next I/P frame, which is under consideration. If the number of backward predicted MBs for the first B-frame ( $N_{B1}$ ) is high (above  $N * T_B$ , where,  $T_B$  - pre-determined percentage threshold to identify an abrupt scene change in B-frames), then most of the MBs for the first B-frame is predicted from the next I/P frame. Therefore this implies that the scene change must have occurred at the first B-frame. If it is below the threshold, then the algorithm checks the number of backward predicted MBs for the second B frame ( $N_{B2}$ ) against the same threshold used for first B- frame. If it is satisfied, then the scene change is declared at second B frame and if it is not satisfied then the scene change should have occurred at next I/P frame. Complete algorithm is shown in Figure 3. It is interesting to note that this algorithm is independent of GOP structure and also it performs in real time.



[Where,  $t$  - Frame number]

Figure 3 : Flowchart for the proposed Algorithm

#### IV. SIMULATION RESULTS

For the simulation, MPEG-2 bit stream (1.152 Mbits/s) is used with 176x144 QCIF format. Figure 4 shows the variation of the number of interpolated MBs as a percentage with respect to the total number of MBs for the first B-frame ( $B_1$ ).

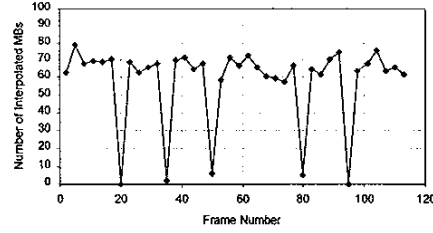


Figure 4 : Number of interpolated MBs (%) with frame number (for first B frame -  $B_1$ )

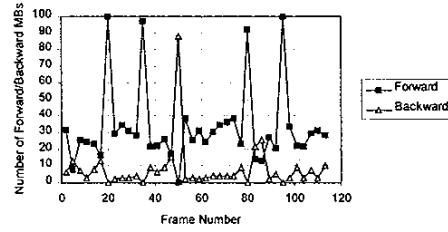


Figure 5 : Number of forward and backward predicted MBs (%) with frame number ( $B_1$ )

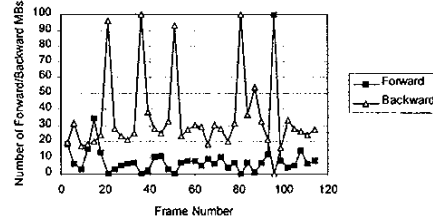


Figure 6 : Number of forward and backward predicted MBs (%) with frame number ( $B_2$ )

When  $\frac{N_{IB}}{N} < T_{IB}$  ( $T_{IB}=15\%$ ), it indicates, the sub-GOP where the abrupt scene changes are. Figure 5 and 6 show the number of forward-predicted MBs and number of backward predicted MBs for first and second B-frames respectively. Now,  $N_{B1}$  and  $N_{B2}$  are compared against  $T_B$  ( $=75\%$ ) and abrupt scene changes are declared accordingly. For instance, the algorithm identifies frame number 20 (where it is below  $T_{IB}$ , Figure 4) as a candidate for sudden scene change

detection. From Figure 5, it is clear that  $N_{B1}$  is less than  $T_B$  and there is no sudden scene change at  $B_1$ . In Figure 6, it is observed that,  $N_{B2}$  exceeds  $T_B$  and hence declares the scene change at  $B_2$  (frame 21). Thus, following the same argument, other scene changes were detected at frame number 36( $B_2$ ), 50( $B_1$ ), 81( $B_2$ ) and 97( $P$ ).

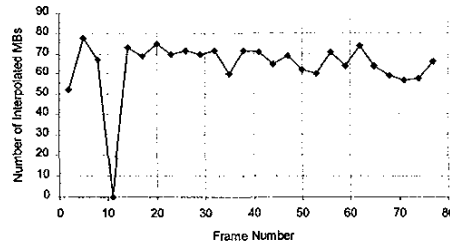


Figure 7 : Number of interpolated MBs (%) with frame number (for a fade-in sequence,  $B_1$ )

Figure 7 shows number of interpolated MBs for the first B-frame with a fade-in sequence. There it shows the exact sub-GOP location where fade-in starts (It is an abrupt scene change). Following the second stage of the algorithm, it is easy to locate the exact scene change within the sub-GOP and it occurred at frame number 11 ( $B_1$ ). If [4] is applied to this sequence, series of false detection will be identified as it is clear from Figure 8 [4]. From Figure 8, it is observed that the number of intra coded MBs for frame number 16(P), 19(P), 22(P) are 93, 57 and 34 respectively. Similarly, the algorithm is checked with other special effects such as fade-out, tilting, panning and observed that the algorithm is capable of detecting all sudden scene changes unlike [4], which produced many false detections.

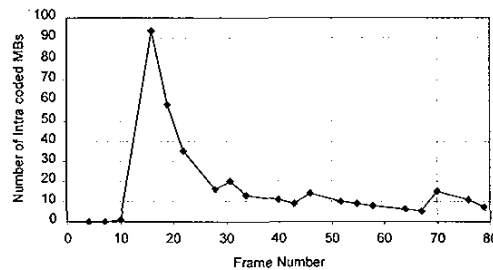


Figure 8 : Number of intra-coded MBs (%) with frame number (for a fade-in sequence, P-frames)

Sequence	Number of Scene Changes	Detected		Undetected		False Detected	
		R	P	R	P	R	P
1	20	15	20	5	0	1	0
2	25	18	25	7	0	38	0

Table 1 : Performance comparison of the proposed algorithm [P] against ratio method [R][4]

Proposed algorithm is tested with two sequences. Sequence 1 has 20 sudden scene changes and it doesn't contain any video special effects. Sequence 2 contains both sudden scene changes and several special effects like fade-in, fade-out, dissolving and panning. This scene is having 25 sudden scene changes altogether. Results presented in table 1, show that the proposed algorithm can detect all sudden scene changes in both sequences and the ratio method couldn't detect sudden scene changes accurately. The performance is worst in sequence 2 as there are large number of false detections in ratio method.

Proposed algorithm is very efficient and eliminates cumbersome variance calculations or partial decompression of the compressed data [2,3,4]. Furthermore, this algorithm can detect abrupt scene changes anywhere in the scene very accurately. This was tested with sequences (which contains more than 100 sudden changes) and was observed to be reliable and accurate.

## V. CONCLUSIONS

This paper presents a real time algorithm, which can detect abrupt scene changes in MPEG-2 compressed video. The scheme used the number of interpolated MBs for a given B-frame as the main feature because it measures how strongly the previous and next I/P frames are correlated. Experimental results show that this algorithm can detect all abrupt scene changes independent of the nature of the sequences and the special effects in the scenes.

## ACKNOWLEDGEMENTS

First author would like to express their gratitude and sincere appreciation to the university of Bristol and CVCP for providing financial support for this work.

## References

- [1]. Sikora T., "MPEG Digital Video-Coding Standards", IEEE Signal Proc. magazine, pp.82-100, Sept. 1997.
- [2]. Yeo, B.L., "Rapid Scene Analysis on Compressed Video", IEEE Trans. on Circuits and Systems for Video Technology, Vol.5, No.6, pp. 533-544, December,1995.
- [3]. Zhang, H.J., " Automatic Partitioning of Full-Motion Video", ACM/Spinger Multimedia Systems, Vol.1, No.1, pp. 10-28, 1993.
- [4]. Shin, T., "Hierarchical scene change detection in an MPEG-2 compressed video sequence", Paper No. 1099, TPA 13-6, ISCAS-1998.
- [5]. Fernando, W.A.C., Canagarajah, C.N., Bull, D.R., "DFD Based Scene Segmentation For H263 Video Sequences", Paper Number-747, 65.21, ISCAS-1999.
- [6]. Fernando, W.A.C., Canagarajah, C.N., Bull, D.R., "Automatic Detection of Fade-in, Fade-out and Dissolving in Video Sequences", Paper Number-748, 74.6, ISCAS-1999.
- [7]. Fernando, W.A.C., Canagarajah, C.N., Bull, D.R., "Video Segmentation and Classification for Content Based Storage and Retrieval Using Motion Vectors", Paper Number-3656-69, SPIE-1999.